From SIGLE to OpenSIGLE and Beyond : An In-depth Look at Resource Migration in the European Context

Joachim Schöpfel, Christiane Stock, Nathalie Henrot

From SIGLE to OpenSIGLE and Beyond: An In-depth Look at Resource Migration in the European Context

Schöpfel, Joachim; Stock, Christiane & Henrot, Nathalie (INIST-CNRS, France)

Abstract:

In 1980, some major European scientific information centres established the "System for Information on Grey Literature in Europe" (SIGLE) to provide access to European grey literature and to improve bibliographic coverage. August 23, 2006, the Luxemburg Register of Commerce and Societies published the liquidation of the association EAGLE that produced the SIGLE database until 2005.

Nevertheless, the former EAGLE member consented to preserve the European co-operation for grey literature and to transform the 1980 model into a sustainable network in the emerging environment of open access to scientific information.

The first step was to archive the SIGLE records in an open and freely searchable database, conform to the OAI metadata harvesting protocol. The French INIST developed OpenSIGLE based on MIT software (DSpace) and loaded most of the SIGLE records in a simplified XML format.

The communication provides an overview of the short history of SIGLE and EAGLE and describes how this unique resource was moved from a traditional host to an open access environment, giving the database a new look while preserving essential features characteristic for SIGLE.

Notes on the authors:

Joachim Schöpfel graduated from the University of Hamburg in 1984. He worked as a research assistant and lecturer at the University of Hamburg, in the Department of Developmental and Educational Psychology, from 1985 to 1990, before obtaining his Ph.D. from the same university in 1992. He is head of the e-publishing and document supply department at INIST-CNRS; at the University of Nancy he taught Culture and Society (1992-2001) and now teaches information sciences (since 2001). He is member of the UK Serials Group and of the international advisory board of COUNTER and was the last president of EAGLE.

Contact address: Dr. Joachim Schöpfel, INIST-CNRS, 2 allée du Parc de Brabois, 54519 Vandoeuvre Cedex, France. E-Mail: joachim.schopfel@inist.fr

Christiane Stock graduated from the University of Freiburg in 1984. She joined INIST-CNRS, the French Institute of Scientific and Technical Information in 1989. Member of the Technical Committee for the SIGLE database from 1993 to 2005, she also set up the national agency for ISRN (International Standard Report Number). Today she is the head of the monographs and grey literature section at INIST.

Contact address: Christiane Stock, INIST-CNRS, 2 allée du Parc de Brabois, 54519 Vandoeuvre Cedex, France. E-mail: christiane.stock@inist.fr

Nathalie Henrot graduated in History, then in Information Sciences from the University of Tours in 1988. She has been working for the INIST-CNRS for seventeen years, more specifically at the Monographs & Grey Literature Section from 1993, for congress proceedings acquisition. She is now the user administrator in the OpenSIGLE project.

Contact address: Nathalie Henrot, INIST-CNRS, 2 allée du Parc de Brabois, 54519 Vandoeuvre Cedex, France

1. EAGLE, a short history

In 1980, some major European scientific information centres established the "System for Information on Grey Literature in Europe" (SIGLE) to provide access to European grey literature and to improve bibliographic coverage (see Wood & Smith 1993).

The SIGLE database covered all scientific domains, STM, social sciences and humanities. In 2005, it contained 855,260 records from 16 countries and the European Commission; UK, German, French and Dutch records represented 90% of the overall input. 63% of the records were reports, 32% were theses and dissertations, and the other records were conference proceedings, data files and translations.

From 1980 to 1985, SIGLE was funded by the Commission of the European Communities (CEC). When CEC financial support ended in 1985, the national centres formed a network for the acquisition, identification and dissemination of grey literature called "European Association for Grey Literature Exploitation" or EAGLE, who became the producer of the SIGLE database.

EAGLE was created as a non-profit making association situated in Luxemburg. In the beginning, membership was limited to the member countries of the European Union (former European Community) but this limitation was abolished in 1995. In the end, 14 countries participated actively, and discussions on partnership were going on with other potential members, especially with East European and North African countries.

The economic model of the association was based on initial funding by the European Commission, from 1985 onwards on membership fees and royalties from servers and products (CD-ROM, records). The SIGLE database was distributed by BLAISE, STN, EI and Ovid (Silverplatter) and in the early nineties by SUNIST in France. Records were also sold to organizations like NERAC for specific use.

Costs were generated by the management of the association and mainly by the operating agent, whose task was to merge the national files into a unique server file every month.

2. Challenges and dead-ends

Twenty years after its creation, EAGLE faced four major challenges:

Internet and new technologies of information and communication: SIGLE offered no solution for online cataloguing, metadata harvesting, links to full-text and other resources.

Organisational structure: EAGLE was unable to take important technical and organisational decisions. Main members resigned from the association or intended to do so.

Coverage: National input became increasingly unrepresentative of the national production of grey literature, and input was continuously decreasing. Electronic documents were not referenced in the database.

Economic model: Investment for the development of the database was not provided.

EAGLE members were aware of these challenges and discussed possible options and solutions. Nevertheless, the organisational structure and the economic model of the non-profit making and low budget association did not allow necessary strategy decisions. It also became obvious then that no institutional member would invest more than its membership fee and that another funding from the EU would be unlikely.

3. Liquidation of EAGLE

In the face of this situation, the 2005 General Assembly from March 14, 2005, held at Karlsruhe, Germany, resolved unanimously to liquidate the association. The President (J. Schöpfel, INIST) and the Vice-President (S. Rehme, FIZ Karlsruhe) were appointed as liquidators.

After satisfaction of claims the remaining funds were distributed on equal parts amongst the members, taking into account the current memberships and the payment of the membership fees during the last years.

The contracts with STN, EI, NERAC and Ovid (SilverPlatter) were cancelled. The EAGLE website was deleted.

All usage rights of the SIGLE database lapsed upon complete liquidation of the association while the copyright on input remained with each member who supplied the records. Nevertheless, the General Assembly asked the operating agent, the FIZ Karlsruhe, for interim conservation of the SIGLE records in XML format beyond the liquidation of SIGLE, for the purpose of archiving and integration into a new European non- profit project. Nearly all of the former EAGLE members signed a declaration of intention regarding this future use of their existing input of the SIGLE database.

The complete liquidation of EAGLE was published by the Luxemburg Register of Commerce and Societies at August 23, 2006.

4. Perspectives and projects beyond EAGLE

At the same General Assembly in 2005, we presented the “MetaGrey Europe” project (see Schöpfel 2006). The objective was to preserve the European co-operation for grey literature and to transform the 1980 model into a sustainable network in the emerging environment of open access to scientific information, especially in the context of the 2003 Berlin Declaration.

The conditions were: no or low budgeting, investment if possible only through direct and non-financial contribution by participating structures (human resources…), no or small-scale organisational structure. EU funding could be helpful (we investigated EU Framework Programmes FP6 and FP7 options) but would generate too much executive work.

The first step was to archive the SIGLE records in an open and freely searchable database, compliant with the OAI metadata harvesting protocol. This part of the project is nearly finished. The French EAGLE member INIST developed OpenSIGLE based on MIT software (DSpace) and loaded most of the SIGLE records in a simplified XML format.

The second step is the identification and developing of European OAI initiatives for grey literature. Most of the former EAGLE members already are involved in such projects, especially for electronic theses and dissertations and reports, or intend to do so (see for instance Boukacem-Zeghmouri & Schöpfel 2006 or Stock, Rocklin & Cordier 2006). What is needed is an inventory of these projects and if necessary, technical and organisational assistance.

The third and most important step will be the development of a gateway to European grey literature. Last year in Nancy, we presented this project as a meta-search engine ("MetaGrey Europe"). The underlying idea was the indexing of European grey collections (catalogues, databases, full-text servers and archives) by a performing search engine that copes with different formats and document types, interdisciplinary and multilinguistic terminology. Whether we can co-operate with an existing search engine (Google, Scirus, Exalead) or need to develop a new search tool, perhaps together with the French-German Quaero project or the German Search Engine Lab at the university of Hannover, needs investigation and discussion.

These two steps will be our challenge for the coming years.

5. More information about the SIGLE database

Before talking about the migration from the host-based database SIGLE to OpenSIGLE, allow us to recall some facts and characteristics concerning SIGLE.

For each SIGLE member country, one or two national structures assumed the acquisition, referencing and document delivery of grey literature, mostly national libraries (UK, Luxembourg) or documentation centres of national research organizations (Italy, France, Germany). Merging of the national files was done by an independent operating agent under contract and the database was hosted on up to three different servers. A CD-ROM was produced from 1992 on by Silverplatter/Ovid.

Some characteristics

Cataloguing rules and classification scheme were adopted from the INIS database produced by the IAEA (International Atomic Energy Agency). The SIGLE classification is derived from the American COSATI scheme. The same scheme is today used by GreySource (GreySource 2006). One important difference to INIS is that SIGLE never included serials (only collections of monographs) and never provided records on an analytical level.

Each national structure sent records in its own language. A search through the entire database was made possible through providing an English translation of the title or English keywords. One of these fields was mandatory. This constraint may seem irrelevant from an American point of view, but for some countries it was a considerable barrier to increase input.

One of the goals of SIGLE from the beginning was to facilitate access to grey documents in Europe:

- Each record contained a clear mention of availability (with or without shelf number).

- Each member country committed to supply the referenced document on demand, whether from its own collections or through interlending service by back-up libraries.

- Useful information on document supply (addresses, conditions) was given on specific help pages or in user-guides.

Contents

SIGLE was started in 1980, but since some members put in older documents, the earliest publications go back to the sixties.

Although SIGLE was a multidisciplinary database, about one third of the records were found in the humanities and social sciences (>35%), followed by medical and life sciences (12%) and physics (9 %). These figures were related to different practices of research communities to publish white only or grey, but also to the willingness of organizations to cooperate. Some of the documents existed only in three copies and it was not always easy for the SIGLE centre to obtain one of them.

Among the document types one finds a majority of reports, followed by doctoral theses or dissertations. In several countries monthly files were obtained by conversion from other catalogues, without a clear identification of the document type, so that the "Miscellaneous" category is a "hold all" for these cases.

Over the years the content of the database improved. For instance, since 1997 English abstracts were added, in particular to Russian records. Members provided English keywords with increasing frequency. Even so, another important project was never realized: the integration of electronic documents, if possible with a link to the full text. Instead members started to build institutional repositories or to provide access to electronic grey literature by other means.

6. OpenSIGLE - moving the resources

Preliminaries

Once INIST decided to continue to make SIGLE available and to allow free access, we looked for the tools.

Throughout frequent visits to institutional or other repositories we had observed that not only non-textual documents became more frequent in these repositories or archives, but also that an increasing number of them provided only bibliographic records, without a link to the full text. This is particularly true for sites where the records come from catalogues, and the full text is added eventually (see Groningen 2006). DOAR and ROAR (see DOAR 2006 and ROAR 2006) provide useful information as to what kind of information awaits you in a given archive or repository. So why not use an open source platform for SIGLE?

INIST had previously adopted the DSpace software for 2 in-house applications: I-Revues, a platform for open access journals and conferences (http://irevues.inist.fr), and LARA, a repository for grey reports, which was presented at GL7 (http://lara.inist.fr).

So we were already familiar with this platform, its advantages and its limits and constraints. Particularly useful in this case, we had acquired a first experience through the two aforementioned applications concerning a mass conversion and mass upload of data into DSpace. FIZ Karlsruhe, the last SIGLE Data Processing Centre (DPC) in office, had provided us with records in different formats, including XML, for evaluation.

A quick comparison of the metadata formats and indexes showed that it should be possible to move the SIGLE data to this software without too great a loss in information details.

The next step was the writing of specifications for the metadata conversion and the customization of DSpace. We also studied closely the old CD-ROM version of SIGLE in order to include a maximum of existing functionalities and indexes as well as online-helps in OpenSIGLE.

We were fortunate to have a student in computer science, specialized in Web applications, A. Adlani, as an intern for 10 weeks. With his help and invaluable suggestions, we were able to realize a prototype close to the final version.

Metadata conversion and migration

DSpace uses a qualified Dublin Core metadata set which is less detailed than the SIGLE metadata we received from the DPC (see DSpace Metadata 2006). FIZ Karlsruhe provided us with records in XML written in the SIGLE format and completed by some server-related fields. Therefore mapping was relatively easy, except for some coded fields and specific information. For example we wanted the English title to appear systematically in the field labelled "Title". In the source record it could be either in the field for the original title or in the field for the English title. Several specific fields from the source format were merged to one field for OpenSIGLE. Others were defined differently to fit with the metadata set. Some qualified fields were added to the metadata set used by DSpace without disturbing the OAI compliance: conference title, report number and availability statement.

The most significant change was a simplification in the document type information. The original SIGLE format distinguished between document type and literature indicator, but diverging conversion practices led to inconsistencies. In OpenSIGLE we propose a simplified list of the principal document types.

For the tests we used the French data and a selection of 1400 records from different countries provided by FIZ Karlsruhe. Several test uploads showed us inconsistencies in the display of data, which we tried to amend as far as possible. Those differences were due to either an "evolution" of some elements over the years or a change of the rules such as the constraint to provide an English title. This is why some recent records showed no information at all in the title column in the results display. Instead it now mentions "No title". Some fields in the source records came in a coded form (language, subject categories and document types). We tried as far as possible to convert them into full wording (e.g. inconsistencies for the language codes). We did not try to amend information which came with typing errors in the input files such as publication dates like 3001.

7. OpenSIGLE on DSpace: new presentation and new look

DSpace allows organizing the contents of a repository according to communities and collections. INIST decided to use 2 types of communities: the member countries and SIGLE subject categories on their primary level. Each country or subject category holds a collection of records. Some minor and less used subject categories were regrouped in one collection. In a mass upload on DSpace each records (or item) can be "attributed" to only one community or collection. We decided to choose the first classification code of each record. Since the files of each member country are treated separately, it is possible to declare also the country community for each record.

Contrary to the CD-ROM version the document type is no longer searchable in OpenSIGLE. We found it interesting to display the information in the list of results, along with the title, the

authors and the publication date. This is not a feature of the basic version of DSpace, but we observed similar practices in other repositories (see ERA 2006 and Glasgow 2006).

The SIGLE classification scheme with its 246 subject sub-categories can be searched through the subject field, either by its code or its wording. A specific help page accessible at any moment lists the complete classification schemes with both the codes and their description. As mentioned above, the subject areas were reduced to 15 entries for the organization of the database in collections and for browsing purposes.

For OpenSIGLE we chose the latest stable version available of the software which was then DSpace 1.3.2. One of the new features in this version is the support of multilingualism of the user interface (cf. DSpace system documentation 2006) This feature has been developed a little bit further by our student and OpenSIGLE can be used with interfaces in English (which is the main version), French, German and Italian, the four most representative languages for the database. The help pages and the "About" information are available in English and French only, since they must be translated specifically.

Document supply being very important for the SIGLE database, we decided to add an order form to facilitate the contact of the holder of the document or ex-EAGLE member. Of course the information about the document availability in each record was maintained, in a specific field. In addition we give updated information for each participating centre on the pages presenting the countries (in the "Countries" collections).

Last, but not least we customized the user interface with the help of our student: look, colours, size etc.

8. OpenSIGLE: results and outlook

The new OpenSIGLE repository gives open access to the bibliographic records of the SIGLE database from the members who signed the agreement.

With the migration to the DSpace platform look and presentation have changed.

Some data like the language or the document type are no longer searchable, but are still displayed, even in the list of results. The principal characteristics of the SIGLE database have been preserved or even improved. Access to the full text will be facilitated through an order form for document delivery and for some records hopefully through links to the electronic version in the future. Since the records are organized in collections based on the subject categories, and the OAI protocol for metadata harvesting considers collections as sets, a selective harvesting by subject will be possible.

More generally, OpenSIGLE seems to be the first migration of an important traditional bibliographic database into an OAI compliant environment. Some factors facilitated this migration: For the mapping of the metadata we had the advantage to pass from a very detailed format to a simpler one. The whole project benefited largely from our previous experience with DSpace and in particular from our knowledge about the import of records. Still OpenSIGLE provided us with a new experience concerning mass uploads on an Open Source platform; it is probably the most important upload for a DSpace repository ever done.

Although no longer updated OpenSIGLE remains a useful source identify and access to European grey literature. Three future developments may improve the visibility and usability of OpenSIGLE:

Links to the full text: Even if the new repository contains only bibliographic records, links from the OpenSIGLE metadata to the electronic full text where available are technically possible but have to be provided by the former EAGLE members.

Completion of content: Hopefully, those of the former EAGLE members who didn't sign the declaration of intention yet may reconsider their position and agree to the import of "their" national SIGLE input into the new database.

Referencing: Linking the OpenSIGLE records to scientific or general search engines, as a part of the MetaGrey Europe project, would largely enhance the visibility of the European grey documents of the last 20 years.

The OpenSIGLE database will be available in the next weeks at the address http://opensigle.inist.fr. Comments and remarks are welcome!

Acknowledgments to Anouar Adlani, a master student in information sciences from the Nancy university, and to Patrick Kremer from the INIST information system department for their helpful ideas and assistance during the preparation and implementation of the OpenSIGLE database.

Bibliography and resources

Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. Berlin 2003. http://oa.mpg.de/openaccess-berlin/berlindeclaration.html

Boukacem-Zeghmouri, C. & Schöpfel, J.: "Document supply and open access: an international survey on grey literature". – In: *Interlending & Document Supply* 2006, 34, 3, 96-104.

DSpace, a digital repository system, [homepage] (2006): http://www.dspace.org/ (accessed November 2006)

DSpace Metadata (2006): Metadata schema used by DSpace http://dspace.org/technology/metadata.html DSpace currently uses a qualified version of the Dublin Core schema based on the Dublin Core Libraries Working Group Application Profile (LAP). Some qualifiers were also added to suit DSpace needs.

DSpace (2006): Dspace System Documentation. http://www.dspace.org/technology/system-docs/history.html Version History/ Indicates as changes for DSpace 1.3 : Initial i18n Support for JSPs - Note: the implementation of this feature required changes to almost all JSP pages (accessed in November 2006).

ERA [homepage] (2006) : Edinburgh Research Archive : http://www.era.lib.ed.ac.uk/index.jsp a digital repository of research output from The University of Edinburgh, especially : http://www.era.lib.ed.ac.uk/browse-title (accessed in November 2006).

Glasgow [homepage] (2006) : Glasgow DSpace Service : https://dspace.gla.ac.uk/index.jsp The Glasgow DSpace Service is a repository for digital content including working papers, technical reports, theses and pre-prints by members of the University of Glasgow. https://dspace.gla.ac.uk/browse-title (last visited November 15th 2006)

GreySource (2006): http://www.greynet.org/greysourceindex.html : GreySource ~ A Selection of Web-based Resources in Grey Literature (accessed in November 2006).

Groningen [homepage] (2006): Dissertations of the University of Groningen http://dissertations.ub.rug.nl (accessed in November 2006).

I-Revues [homepage] (2006): http://irevues.inist.fr (accessed in November 2006). Website providing access to full text of serials and conferences.

LARA – Libre Accès aux Rapports [homepage] (2006): http://www.lara.inist.fr A federative repository for grey reports (accessed in November 2006).

OpenDOAR [homepage] (2006): The Directory of Open Access Repositories - OpenDOAR http://www.opendoar.org/index.html OpenDOAR is an authoritative directory of academic open access repositories (accessed in November 2006)

ROAR [homepage] (2006): Registry of Open Access Repositories (ROAR) http://archives.eprints.org/index.php (accessed in November 2006)

Schöpfel, J.: "MetaGrey Europe, A Proposal in the Aftermath of EAGLE-SIGLE". – In: Farace, D. & Frantzen, J. (ed.): *GL7 Conference Proceedings. Seventh International Conference on Grey Literature: Open Access to Grey Resources.* Nancy, 5-6 December 2005. - Amsterdam: TextRelease, 2006.

Stock, C., Rocklin, E. & Cordier, A.: "LARA – open access to scientific and technical reports". – In: Farace, D. & Frantzen, J. (ed.): *GL7 Conference Proceedings. Seventh International Conference on Grey Literature: Open Access to Grey Resources.* Nancy, 5-6 December 2005. - Amsterdam: TextRelease, 2006.

Wood, D.N. & Smith, A.W.: "SIGLE: A Model for International Co-operation". – In: *Interlending & Document Supply* 1993, 21, 1, 18-22.